Physics-Informed Diffusion Models for Unsupervised Anomaly Detection in Multivariate Time Series

Juhi Soni[1], Markus Lange-Hegermann[2], and Stefan Windmann[1]

[1] Fraunhofer IOSB-INA, Lemgo, Germany
{juhi.soni,stefan.windmann}@iosb-ina.fraunhofer.de
[2] TH-OWL University of Applied Sciences and Arts Lemgo, Germany
markus.lange-hegermann@th-owl.de

Abstract. We propose an unsupervised anomaly detection approach based on a physics-informed diffusion model for multivariate time series data. Over the past years, diffusion model has demonstrated its effectiveness in forecasting, imputation, generation, and anomaly detection in the time series domain. In this paper, we present a new approach for learning the physics-dependent temporal distribution of multivariate time series data using a weighted physics-informed loss during diffusion model training. A weighted physics-informed loss is constructed using a static weight schedule. This approach enables a diffusion model to accurately approximate underlying data distribution, which can influence the unsupervised anomaly detection performance. Our experiments on synthetic and real-world datasets show that physics-informed training improves the F1 score in anomaly detection; it generates better data diversity and log-likelihood. Our model outperforms baseline approaches, additionally, it surpasses prior physics-informed work and purely data-driven diffusion models on a synthetic dataset and one real-world dataset while remaining competitive on others.

Keywords: Time Series · Anomaly Detection · Physics-informed machine learning · Diffusion Model.

1 Introduction

Anomaly detection techniques play an important role in various applications. They can be used to identify system defects, failures, network attacks, and more. In the industry, detecting abnormal patterns in data is crucial for ensuring smooth operations and reducing process downtime [29]. Researchers have introduced many supervised, semi-supervised, and unsupervised anomaly detection techniques over the years, driven by the increasing availability of data. Given the limited requirement for labeled data in semi-supervised methods and the absence of such a requirement in unsupervised methods, research efforts have concentrated on semi-supervised and unsupervised learning approaches [11,29]. The generative models have shown their capability to generate high-quality samples

across various domains [9,12,18,26,35], such as text-to-image generation, audio generation, video generation, etc. In recent years, there has been a growing interest in the application of generative models in time series analysis, particularly for anomaly detection. These models [9,12,18,35] have been applied to a range of time series tasks, including forecasting, imputation, synthetic data generation, and anomaly detection. Generative models are used in unsupervised anomaly detection because of their ability to model complex data distributions by learning the underlying structure of normal data and identify anomalies as deviations from the expected patterns. In time series anomaly detection, various variants of Variational Autoencoders (VAEs) [10,21,36,40], Generative Adversarial Networks (GANs) [16, 20, 28], diffusion models [14, 29, 37, 38], and auto-regressive models are commonly used.

In the last few years, the diffusion model has gained popularity due to its ability to generate high-quality samples, optimize log-likelihood, mode coverage, and training stability [12,17,32,35,39]. In the time series domain, diffusion model based reconstruction and imputation techniques for anomaly detection are introduced in [14, 29, 37, 42]. All these anomaly detection approaches rely solely on the data. In domains such as fluid dynamics, biology, and engineering, data inherently adhere to physical laws, and the standard diffusion model training and sampling ignore these underlying physical laws. Therefore, to incorporate physics into the model, in [33], physics residual is included as conditioning information for high-fidelity flow field reconstruction. Bastek et al. [2] defined a physics-informed loss function using Partial Differential Equation (PDE) and scaled variance of the diffusion process, with a focus on minimizing the physics constraints during model training. All previous works on physics-informed diffusion models [2,6,15,33] are focused on applying PDE constraints during model training or sampling on image datasets.

In the time series domain, the generative model must capture the temporal relationships in addition to the feature distribution [8]. These temporal relationships in some datasets follow the laws of physics. Incorporating prior knowledge improves the model's fit and enhances its interpretability [3]. Previous work on diffusion models for time-series anomaly detection has not exploited physical laws during training or inference. Inspired by this recent work on anomaly detection introduced above [14, 29, 37, 42] and the physics-informed diffusion model [2, 33], we propose a temporal physics-informed diffusion model for multi-variant time series anomaly detection. Here, we aim to improve anomaly detection capabilities by exploiting the physical knowledge of the data during model training.

In this paper, we propose a novel approach to add physics information during model training. We define a weight for every diffusion time step and use it to build the physics-informed loss function for physics aware diffusion model training. We design a weighting schedule that reduces the influence of noisy data relative to less noisy data in the construction of the physics-informed loss function. Our approach improves the model's ability to learn the underlying temporal distribution of data, ultimately resulting in improved robustness for anomaly detection. In this work, we have evaluated the proposed anomaly detection method

on four multivariate time series datasets in terms of F1 score, log-likelihood, and data diversity. To validate our approach, we assessed the statistical significance of the performance gains achieved through physics-informed training.

Our contributions are summarized as follows:

1. We develop a temporal physics-informed diffusion model for anomaly detection that adheres to the underlying physical laws governing the data by incorporating weighted physics-informed loss function during diffusion model training.
2. We define and incorporate a weight schedule into the physics-informed loss function.
3. Assessing the physics-informed and uninformed diffusion model by validating them based on the log-likelihood, data diversity, and F1 score.

2   Background

2.1   Evidence Lower Bound (ELBO)

In unsupervised anomaly detection, we can model the normal behavior of data using a probability distribution by leveraging the Evidence Lower Bound (ELBO) objective, and then define a threshold to distinguish between normal and anomalous data. The ELBO is a lower bound on the marginal likelihood of data x and a good approximation to $\log p(x)$. Within a variational inference scheme, we introduce latent variables, z, sampled from a prior distribution $p(z)$. A simpler distribution, $q(z|x)$, is then used to approximate the true posterior, $p(z|x)$ [4,26]. These posterior approximations enable the model to capture the underlying distribution of normal data, facilitating the more effective identification of deviations or anomalies in the latent space.

$$\mathbf{ELBO} = \mathbb{E}[\log p(x|z)] - D_{KL}(q(z|x)||p(z|x))$$

So, by maximizing the ELBO, we ensure the variational distribution $q(z|x)$ closely approximates the true posterior $p(z|x)$ by minimizing their Kullback-Leibler D_{KL} divergence.

2.2   Denoising Diffusion Probabilistic Models (DDPM)

A diffusion probabilistic model is a latent variable model parameterized by a Markov chain, which is trained using variational inference techniques to generate samples that closely match the data within a finite time frame [12]. The training of this model consists of two phases: a forward diffusion process and a reverse denoising process. In the forward process, the approximate posterior is fixed to a Markov chain that gradually introduces Gaussian noise to the data following a variance schedule $\sigma_1, ..., \sigma_T$ [12]. The forward process is defined by

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_0, (1 - \alpha_t)I)$$

where $\alpha_t = 1 - \sigma_t$.

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\overline{\alpha_t}}x_0, (1 - \overline{\alpha_t})I)$$

where $\overline{\alpha_t} = \prod_{s=1}^{t} \alpha_s$. The reverse diffusion process is a generative process in which we invert the forward diffusion process. It is defined by

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

where $\Sigma_\theta(x_t, t) = \sigma_t^2 I$ or $\frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\sigma_t$ [12] and the generator $\mu_\theta(x_t, t)$ approximate the data distribution. The combination of forward process q and reverse process p constitutes a variational autoencoder [18, 19, 27], and the model is trained by maximizing the ELBO.

$$\mathcal{L}_{\mathcal{DM}} = \underbrace{D_{KL}(q(x_T|x_0)||p(x_T))}_{L_T} + \underbrace{\sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}_{L_{t-1}}$$
$$- \underbrace{\log p_\theta(x_0|x_1)}_{L_0}$$

The L_T and L_0 computed using standard techniques [19] and since all distributions are Gaussian L_{t-1} can be defined by [12]

$$L_{t-1} = \mathbb{E}_q \left[\frac{1}{2\sigma^2}||\tilde{\mu}(x_t, x_0) - \mu_\theta(x_t, t)||^2 \right] + C$$

As per [12], the model can be parameterized as a noise ϵ prediction model ϵ_θ by using the simplified ELBO.

$$\mathcal{L}_{\mathcal{DM}} = \mathbb{E}_{t,x_0,\epsilon} \left[||\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha_t}}x_0 + \sqrt{1 - \overline{\alpha_t}}\epsilon, t)||^2 \right]$$

In addition, the neural network x_θ can also be parameterized to predict clean data x_0.

2.3   Physics-informed Neural Network

Physics-informed Neural Network (PINN) [30, 31] works as a data-driven function approximator, where physical laws are exploited in the training process. Many laws of physics are described by Partial Differential Equations (PDE) or Ordinary Differential Equations (ODE). In Physics-informed Neural Networks, we train a model with the underlying physics knowledge of data by embedding this knowledge into the network's loss function [7, 22, 23, 30]. The overall training objective is a composite loss function, $\mathcal{L}$, which combines a standard network loss, $\mathcal{L}_\mathcal{F}$, with a physics-informed loss $\mathcal{L}_{\mathcal{PI}}$.

$$\mathcal{L} = \mathcal{L}_\mathcal{F} + \mathcal{L}_{\mathcal{PI}}$$

The physics-informed loss, $\mathcal{L}_{\mathcal{PI}}$ is defined by

$$\mathcal{L}_{\mathcal{PI}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}(\mathcal{D}, F(x_i; \theta))^2$$

where $\mathcal{E}$ denotes the residual of the governing physical equations, $\mathcal{D}$ denotes the physics prior, and $F(x; \theta)$ represents the output of neural network. The physics-informed loss term works as a regularizer that limits model parameters to applied physics law. Encoding physics information into a learning algorithm enhances the informational content of the data, leading to a boost in neural network model performance [30].

3   Related Work

The physics-aware diffusion model was introduced in [15] to generate samples constrained by partial differential equations (PDE) by minimizing the residue during model sampling. Christopher et al. [6], augmented diffusion-based synthesis with constraints during sampling to generate high-fidelity content. A physics-guided motion diffusion model (PhysDiff) [44] incorporates physical constraints into the diffusion process to generate physically implausible motions. For high-fidelity flow field reconstruction in [33], physics residual is included as conditioning information during model training and inference. Similar to our approach, [2] introduces a physics-informed residual based on partial differential equations (PDEs) and the scaled variance of the diffusion process. This residual is incorporated into the training of a diffusion model to generate samples that satisfy the underlying PDE. These studies focus on incorporating physics-based constraints into model training or sampling to produce physically consistent samples. Apart from [2] and [44], they overlook the noise components intrinsic to the diffusion process. Our weighting schedule mitigates the impact of noise during physics-informed training, resulting in a closer fit to the data distribution.

In time series anomaly detection in [29] the autoencoder model is used with the diffusion model to identify anomalies using a reconstruction signal. A Denoising Diffusion Time Series Anomaly Detection (DDTAD) [37] also utilizes the reconstruction error for detecting abnormalities. Xiao et al. [42] have used a generative time series imputation process for anomaly detection. A TimeADDM [14] uses multi-level reconstructions of diffusion models for unsupervised anomaly detection for multivariant time series. The ImDiffusion [5] method reduces the uncertainty of data and enhances anomaly detection by employing imputation to integrate information from neighboring time points, effectively capturing temporal and correlated patterns. Analogous to our approach in [45] normal data distribution features are learned during training and abnormal data are detected through sequence reconstruction. The benefit of our work over existing methods lies in the utilization of ELBO for the detection of anomalous patterns, and the integration of a weighted physics law during model training yields a physically

consistent model, resulting in a statistically significant improvement in anomaly detection performance.

4   Temporal Physics-informed Diffusion Model (TPIDM)
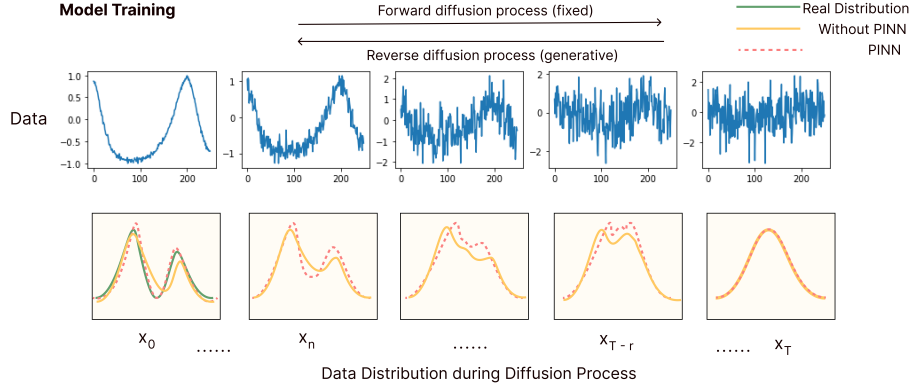
Fig. 1. Data-driven diffusion model vs physics-informed diffusion model training. A weighted Physics-informed loss is added to the diffusion model training, which leads to an improvement in the model's ability to approximate the underlying data distribution.

To learn the physics-dependent temporal distribution of data, the underlying physics of the data is integrated into the loss function of the diffusion model, thereby embedding the physical knowledge of the data into the model's learning process. The physics-informed loss is formulated using the reconstructed signal and the underlying physical laws governing the data. It is evident from Figure 1 that the data near the time step $t = 0$ exhibits significantly lower noise levels, indicating that earlier diffusion steps preserve more of the original signal structure. Furthermore, at earlier diffusion steps, the signal is closer to being a smooth curve; hence, derivatives can be estimated much more reliably. In contrast, when reconstructing the signals $\widehat{x_0}$ from x_t at diffusion time step T or near T, the resulting outputs contains higher noise levels than those reconstructed from earlier time steps, as the model has not yet reached convergence. Due to these noisy reconstructed signals, the physics-informed loss is inadequately defined because of the highly noisy derivative estimations. This adversely affects the model's learning process, causing divergence from optimal solutions and resulting in anomalous or improbable outputs. So, to address this poorly defined physics-informed loss and enhance the model's performance, we propose a static weight schedule for physics-informed loss functions. The physics-informed loss weight schedule, denoted $\overline{\lambda_{PI_t}}$, is constructed to prioritizes toward the initial n

diffusion steps by assigning them greater weight in the loss computation, while subsequent steps approaching the final step T are assigned zero weight.

$$\mathcal{L}_{\mathcal{PI}} = \overline{\lambda_{PI_t}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}(\mathcal{D}, \widehat{x_{i_0}})^2$$

Here $\overline{\lambda_{PI_t}} = \prod_{s=1}^{t} \lambda_{PI_s}$ and λ_{PI_s} is defined by $f(s) \cdot (m - n) + l$, where $f(s)$ is a nonlinear function, m and n determine the amplitude bounds, l is offset. Here $\overline{\lambda_{PI_t}}$ eliminates the noise in the reconstructed signal, $\widehat{x_{i_0}}$. This noise which is inherited from signal x_{i_t} at an arbitrary diffusion step t, and its removal ultimately enhances the reliability of the $\mathcal{L}_{\mathcal{PI}}$ computation. This weight schedule is bounded between 0 and 1, and it saturates at 0. Its shape and saturation points depend on the $f(s)$. This weighting schedule ensures that the constructed physics loss is more precise than one computed using all diffusion steps. The Diffusion Model loss function with physics-informed loss is as follows

$$\mathcal{L}_{\mathcal{TPIDM}} = \mathcal{L}_{DM} + \mathcal{L}_{\mathcal{PI}}$$

where $\mathcal{L}_{DM}$ indicates the diffusion model loss function.

4.1   Anomaly Detection
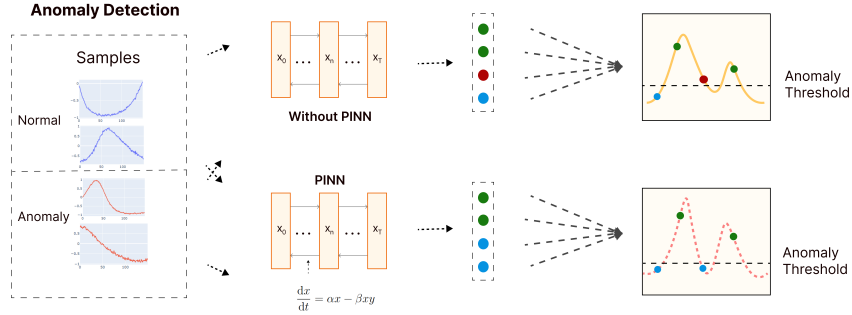
Fig. 2. Physics-informed training influences the model's ability to learn and generalize the underlying data distribution, resulting in the correct identification of normal and anomaly data compared to an uninformed diffusion model.

The diffusion model learns to approximate the underlying probability distribution of the data. After training the model exclusively on normal data, anomaly detection is performed based on its Evidence Lower Bound (ELBO), which serves as the diffusion model's loss function $\mathcal{L}_{DM}$. As shown in Figure 2, a testing sample is labeled as an anomaly if the computed ELBO of the sample exceeds the

threshold. The anomaly threshold is determined using [37]

$$\mathcal{A}_{thr} = \mu_{trim} + k \cdot \text{iqr}$$

with μ_{trim} and iqr denoting the trimmed mean and interquantile range of the model's loss over the validation data, respectively.

5   Experiments

The proposed anomaly detection method has been evaluated using four multivariate time series datasets (three real-world datasets and a synthetic dataset described in Appendix A). The TPIDM's anomaly detection results are compared against vanilla autoencoder, Variational Autoencoder, K-means algorithm, and two Physics-informed [2, 33] diffusion models. The diffusion models are further evaluated quantitatively using log-likelihood estimates on the validation dataset and the binary F1 score to assess anomaly classification performance. We use imbalanced data (700 normal and 300 anomalies) to evaluate the performance of a model in anomaly detection. We investigate the statistical significance of performance improvements achieved by the physics-informed diffusion model over the uninformed diffusion model and baseline methods for anomaly detection. The diversity of the generated data is evaluated by the Principle Component Analysis (PCA) [25] charts of the original and synthetic data [8, 43]. Details regarding the model architecture, hyperparameters, model training, and inference time are provided in Appendix B.

5.1   ELBO and Sample Diversity

Table 1 shows the impact of our method on the likelihood. We achieved better log-likelihoods across the datasets EMPS, Predator-Prey, and Air Compressor for a noise prediction model (ϵ_θ). The x_0 prediction model outperformed the other methods on the EMPS, Lenze, and Air Compressor datasets. Prior physics-informed [2, 33] diffusion models' results are compared based on their diffusion loss (ϵ or x_0). Figure 3 illustrates the performance of the models with respect to data diversity. It shows that our method improved the diversity and performed better compared to a data-driven model and the model proposed in [33] when trained via noise prediction objective. However, training the model (x_θ) using the x_0 prediction objective did not yield any improvement in sample diversity.

Table 1. ELBO comparison of data-driven diffusion model (DM without PINN), prior physics-informed diffusion model introduced in [2, 33] (PIDM), and our method.

	ELBO			
	EMPS	**Predator-Prey**	**Lenze**	**Air Compressor**
DM (ϵ_θ)	-703.56 ± 0.45	-612.46 ± 0.12	-4779.47 ± 3.13	-4224.99 ± 2.47
PIDM [33]	-628.46 ± 2.02	-591.85 ± 0.86	**-4816.08 ± 1.26**	-3993.55 ± 11.57
TPIDM (Ours, ϵ_θ)	**-707.99 ± 0.81**	**-613.15 ± 0.00**	-4767.04 ± 3.39	**-4301.11 ± 4.34**
DM (x_θ)	-819.21 ± 0.69	**-490.87 ± 0.00**	-6291.91 ± 0.00	-4061.78 ± 0.00
PIDM [2]	28989.58 ± 3075.11	708.47 ± 0.04	-5289.29 ± 0.01	-761.68 ± 121.41
TPIDM (Ours, x_θ)	**-821.78 ± 0.02**	-469.81 ± 0.00	**-6303.48 ± 0.00**	**-4111.57 ± 2.02**

(a) Diffusion model without PINN

(b) PIDM [33]

(c) TPIDM (ours)
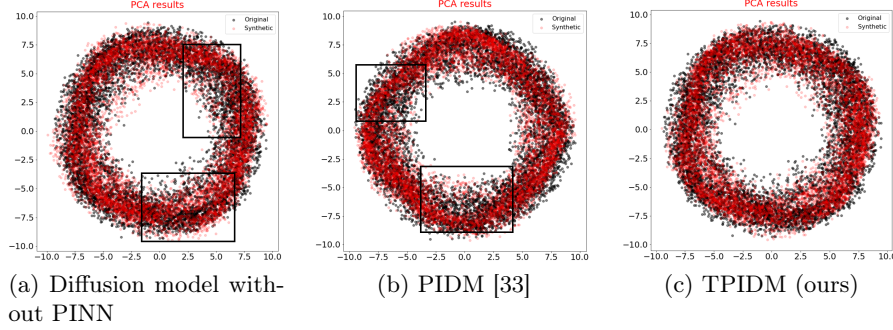
Fig. 3. PCA plots of generated and original Lenze data. The red portion represents the synthetic data, while the black portion corresponds to the original data. The bounding boxes highlight areas not covered by the generated data in Figures 3a and 3b. These plots indicate that our model achieves more overlap between the red and black regions between the synthetic and original data, suggesting improved data diversity.

5.2 Anomaly Detection Results

Table 2. F1 score comparison of AE, VAE, K-means, data-driven diffusion model (without PINN), prior physics-informed diffusion model [2, 33], and our method

	F1 Score			
	EMPS	**Predator-Prey**	**Lenze**	**Air Compressor**
Other models				
VAE	0.9624 ± 0.0047	0.9133 ± 0.0011	0.8298 ± 0.0031	0.9664 ± 0.0389
AE	0.9168 ± 0.0030	0.8226 ± 0.0000	0.9820 ± 0.0000	0.9848 ± 0.0000
K-MEANS	0.8526 ± 0.0000	0.7868 ± 0.0000	0.8102 ± 0.0000	0.8235 ± 0.0000
DM with ϵ_θ				
DM (ϵ_θ)	0.9627 ± 0.0031	0.9712 ± 0.0017	$\mathbf{0.9923 \pm 0.0003}$	1.0 ± 0.0000
PIDM [33]	0.9606 ± 0.0067	0.9586 ± 0.0026	0.9405 ± 0.0167	1.0 ± 0.0000
TPIDM (Ours, ϵ_θ)	$\mathbf{0.9699 \pm 0.0028}$	$\mathbf{0.9830 \pm 0.0000}$	0.9622 ± 0.0116	1.0 ± 0.0000
DM with x_θ				
DM (x_θ)	$\mathbf{1.0 \pm 0.0000}$	0.9957 ± 0.0000	$\mathbf{0.9993 \pm 0.0000}$	1.0 ± 0.0000
PIDM [2]	0.85 ± 0.0014	$\mathbf{1.0 \pm 0.0000}$	0.9624 ± 0.0000	1.0 ± 0.0000
TPIDM (Ours, x_θ)	$\mathbf{1.0 \pm 0.0000}$	$\mathbf{1.0 \pm 0.0000}$	$\mathbf{0.9993 \pm 0.0000}$	1.0 ± 0.0000

Table 2 shows the effectiveness of our method in detecting anomalies. It demonstrates that our method improves the F1 score and outperforms other methods for EMPS and Predator-Prey datasets when a ϵ prediction model (ϵ_θ) is used for training. The Table 2 also reveals that the proposed physics-informed training approach results in higher F1 scores than the models proposed in previous work [2, 33]. Our method also improves the F1 score on the Predatory-Prey dataset when training an x_0 prediction model (x_θ). The x_θ model demonstrates competitive performance against DM and PIDM [2] models as a result of excluding highly noisy components from the PINN loss, while the DM loss remains fixed

throughout the training process, thereby limiting its impact. These results indicate that physics-informed training leads to more accurate models for anomaly detection. Furthermore, these results show that the diffusion model performs better in anomaly detection compared to the other models. Our results demonstrate a statistically significant effect of the physics-informed training in anomaly detection, with a p-value less than 0.01 with the Wilcoxon test when compared with prior work and a purely data-driven approach with ten repetitions.

5.2.1   Ablations

Table 3. F1 score for different physics-informed schedules: The rows correspond to the F1 score of different PINN weight schedules. These numbers show the superior and competitive performance of physics-informed training in anomaly detection.

	F1 Score			
	EMPS	**Predator-Prey**	**Lenze**	**Air Compressor**
DM (ϵ_θ)	0.9627 ± 0.0031	0.9712 ± 0.0017	$\mathbf{0.9923 \pm 0.0003}$	1.0 ± 0.0000
Log-Sigmoid (ϵ_θ)	$\mathbf{0.9699 \pm 0.0028}$	0.9738 ± 0.0017	0.9622 ± 0.0116	1.0 ± 0.0000
Hard-Sigmoid (ϵ_θ)	0.9697 ± 0.0025	0.9552 ± 0.0027	0.9578 ± 0.0152	1.0 ± 0.0000
Sigmoid (ϵ_θ)	0.9447 ± 0.0082	$\mathbf{0.9830 \pm 0.0000}$	0.9104 ± 0.0115	1.0 ± 0.0000
ReLU (ϵ_θ)	0.9015 ± 0.0000	0.9749 ± 0.0017	0.9678 ± 0.0089	1.0 ± 0.0000
DM (x_θ)	1.0 ± 0.0000	0.9957 ± 0.0000	$\mathbf{0.9993 \pm 0.0000}$	1.0 ± 0.0000
Log-Sigmoid (x_θ)	1.0 ± 0.0000	0.9887 ± 0.0000	$0.9985 \pm 4.71e^{-7}$	1.0 ± 0.0000
Hard-Sigmoid (x_θ)	1.0 ± 0.0000	0.9971 ± 0.0000	0.9979 ± 0.0000	1.0 ± 0.0000
Sigmoid (x_θ)	1.0 ± 0.0000	$\mathbf{1.0 \pm 0.0000}$	0.9986 ± 0.0000	1.0 ± 0.0000
ReLU (x_θ)	1.0 ± 0.0000	0.9971 ± 0.0000	$\mathbf{0.9993 \pm 0.0000}$	1.0 ± 0.0000

We investigated the effect of various physics-informed weight schedules on model performance. The anomaly detection results for different schedules are shown in Table 3. The results shown in the table demonstrate that the physics-informed model trained with a logarithmic sigmoid schedule exhibits superior overall performance in comparison to other PINN weight schedules for a ϵ_θ diffusion model. For the Lenze dataset, we did not find the physics-informed training beneficial in the context of the anomaly detection with (ϵ_θ) model. We also explored a few other PINN weight schedules to improve the results for Lenze datasets but did not find them useful. We tried sine and tanhshrink functions in building the PINN weight schedule and trained the diffusion model with these schedules, which improved the F1 score to 0.98. However, these functions did not outperform the other PINN schedules and, in fact, led to a deterioration in performance, reducing the F1 score to 0.94 on the EMPS dataset. On the other hand, the ReLU schedule overall performs competitively compared to other schedules for a x_θ model, and a sigmoid schedule improved the model's performance in identifying anomalies for the Predator-Prey dataset.

6   Conclusion

We introduced a novel approach to learn the physics-dependent temporal relationship of data using a weighted physics-informed loss in diffusion model

training. In particular, we developed a PINN weighting schedule and embedded it within the physics-informed loss formulation. We demonstrated the effectiveness of the proposed method in terms of unsupervised anomaly detection, log-likelihood, and data diversity. Our experimental results indicate that physics-informed training results in an improved F1 score for anomaly detection. Our findings further confirm that with physics-informed training, we achieve better estimation of the data distribution, leading to an improved log-likelihood. Optimizing with the noise prediction objective enhances the diversity of generated samples compared to the purely data-driven model and prior physics-informed work. It shows that our approach performs better than the previous methods by using a novel weighted physics-informed loss scheduled across diffusion steps. Our approach broadly applies to time series governed by physical principles and can be extended to domains such as fluid dynamics and medical imaging, where physical constraints are relevant. One limitation of the model is its longer inference time compared to methods such as AE and VAE, due to its dependence on the number of diffusion time steps.

References

1. A. Janot, M.G., Brunot, M.: Data set and reference models of emps. In: Nonlinear System Identification Benchmarks (2019)
2. Bastek, J.H., Sun, W., Kochmann, D.M.: Physics-informed diffusion models. arXiv preprint arXiv:2403.14404 (2024)
3. Besginow, A., Lange-Hegermann, M.: Constraining gaussian processes to systems of linear ordinary differential equations. Advances in Neural Information Processing Systems **35**, 29386–29399 (2022)
4. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: A review for statisticians. Journal of the American statistical Association **112**(518), 859–877 (2017)
5. Chen, Y., Zhang, C., Ma, M., Liu, Y., Ding, R., Li, B., He, S., Rajmohan, S., Lin, Q., Zhang, D.: Imdiffusion: Imputed diffusion models for multivariate time series anomaly detection. arXiv preprint arXiv:2307.00754 (2023)
6. Christopher, J.K., Baek, S., Fioretto, F.: Projected generative diffusion models for constraint satisfaction. arXiv preprint arXiv:2402.03559 (2024)
7. Cuomo, S., di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F.: Scientific machine learning through physics-informed neural networks: Where we are and what's next (2022)
8. Desai, A., Freeman, C., Wang, Z., Beaver, I.: Timevae: A variational auto-encoder for multivariate time series generation. arXiv preprint arXiv:2111.08095 (2021)
9. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks (2014)
10. Guo, Y., Liao, W., Wang, Q., Yu, L., Ji, T., Li, P.: Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach. In: Asian Conference on Machine Learning. pp. 97–112. PMLR (2018)
11. Hammerbacher, T., Lange-Hegermann, M., Platz, G.: Including sparse production knowledge into variational autoencoders to increase anomaly detection reliability. In: 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE). pp. 1262–1267. IEEE (2021)

12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
13. Hoppensteadt, F.: Predator-prey model. Scholarpedia **1**(10), 1563 (2006)
14. Hu, R., Yuan, X., Qiao, Y., Zhang, B., Zhao, P.: Unsupervised anomaly detection for multivariate time series using diffusion model. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 9606–9610. IEEE (2024)
15. Jacobsen, C., Zhuang, Y., Duraisamy, K.: Cocogen: Physically-consistent and conditioned score-based generative models for forward and inverse problems. arXiv preprint arXiv:2312.10527 (2023)
16. Jiang, W., Hong, Y., Zhou, B., He, X., Cheng, C.: A gan-based anomaly detection approach for imbalanced industrial time series. IEEE Access **7**, 143608–143619 (2019)
17. Kingma, D., Salimans, T., Poole, B., Ho, J.: Variational diffusion models. Advances in neural information processing systems **34**, 21696–21707 (2021)
18. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2022)
19. Kingma, D.P., et al.: Variational inference & deep learning: A new synthesis (2017)
20. Li, D., Chen, D., Jin, B., Shi, L., Goh, J., Ng, S.K.: Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In: International conference on artificial neural networks. pp. 703–716. Springer (2019)
21. Lin, S., Clark, R., Birke, R., Schönborn, S., Trigoni, N., Roberts, S.: Anomaly detection for time series using vae-lstm hybrid model. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4322–4326. Ieee (2020)
22. Meng, C., Griesemer, S., Cao, D., Seo, S., Liu, Y.: When physics meets machine learning: A survey of physics-informed machine learning. Machine Learning for Computational Science and Engineering **1**(1), 1–23 (2025)
23. Misyris, G.S., Venzke, A., Chatzivasileiadis, S.: Physics-informed neural networks for power systems. In: 2020 IEEE Power & Energy Society General Meeting (PESGM). pp. 1–5. IEEE (2020)
24. Mueller, P.N.: Attention-enhanced conditional-diffusion-based data synthesis for data augmentation in machine fault diagnosis. Engineering Applications of Artificial Intelligence **131**, 107696 (2024). https://doi.org/https://doi.org/10.1016/j.engappai.2023.107696
25. Murphy, K.P.: Probabilistic Machine Learning: An introduction. MIT Press (2022), probml.ai
26. Murphy, K.P.: Probabilistic Machine Learning: Advanced Topics. MIT Press (2023), http://probml.github.io/book2
27. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. pp. 8162–8171. PMLR (2021)
28. Niu, Z., Yu, K., Wu, X.: Lstm-based vae-gan for time-series anomaly detection. Sensors **20**(13), 3738 (2020)
29. Pintilie, I., Manolache, A., Brad, F.: Time series anomaly detection using diffusion-based models. In: 2023 IEEE International Conference on Data Mining Workshops (ICDMW). pp. 570–578. IEEE (2023)
30. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv preprint arXiv:1711.10561 (2017)
31. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics informed deep learning (part ii): Data-driven discovery of nonlinear partial differential equations. arXiv preprint arXiv:1711.10566 (2017)

32. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
33. Shu, D., Li, Z., Farimani, A.B.: A physics-informed diffusion model for high-fidelity flow field reconstruction. Journal of Computational Physics **478**, 111972 (2023)
34. Sidebotham, G.: Compressors and the ideal gas. In: An Inductive Approach to Engineering Thermodynamics, pp. 301–353. Springer (2022)
35. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
36. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2828–2837 (2019)
37. Sui, J., Yu, J., Song, Y., Zhang, J.: Anomaly detection for telemetry time series using a denoising diffusion probabilistic model. IEEE Sensors Journal (2024)
38. Tian, F., Shi, X., Zhou, L., Chen, L., Ma, C., Zhu, W.: Prodiffad: Progressively distilled diffusion models for multivariate time series anomaly detection in joint-cloud environment. In: 2024 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2024)
39. Vahdat, A., Kreis, K., Kautz, J.: Score-based generative modeling in latent space. Advances in neural information processing systems **34**, 11287–11302 (2021)
40. Windmann, S., Westerhold, T.: Fault detection in automated production systems based on a long short-term memory autoencoder. at-Automatisierungstechnik **72**(1), 47–58 (2024)
41. Wißbrock, P., Müller, P.N.: Lenze motor bearing fault dataset (lenze-mb) (2025). https://doi.org/10.5281/zenodo.14762423, https://doi.org/10.5281/zenodo.14762423
42. Xiao, C., Gou, Z., Tai, W., Zhang, K., Zhou, F.: Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 2742–2751 (2023)
43. Yoon, J., Jarrett, D., Van der Schaar, M.: Time-series generative adversarial networks. Advances in neural information processing systems **32** (2019)
44. Yuan, Y., Song, J., Iqbal, U., Vahdat, A., Kautz, J.: Physdiff: Physics-guided human motion diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16010–16021 (2023)
45. Zuo, H., Zhu, A., Zhu, Y., Liao, Y., Li, S., Chen, Y.: Unsupervised diffusion based anomaly detection for time series. Applied Intelligence **54**(19), 8968–8981 (2024)

A   Data Sets

A.1   Lotka–Voterra Predator-Prey Dataset

We generate 100,000 samples of a Lotka–Volterra predator-prey [13] model with an initial value (x, y) = (10, 2), where x and y denote the population size of preys and predators, respectively. This non-linear biological model describes the interaction between the two species *Predator* and *Prey*. This data set has two features; one describes the growth rate of Prey, and the other describes the growth rate of Predator.

Prey Growth Rate	**Predator Growth Rate**

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha x - \beta xy \qquad\qquad \frac{\mathrm{d}y}{\mathrm{d}t} = \delta xy - \gamma y$$

Here α describes the maximum growth of prey and β describes the effect of the presence of predator on the growth rate of prey. A δ describes the death rate of predator and γ describes the effect of the presence of prey on the growth of predator. The values of α, β, δ, and γ are 1.1, 0.4, 0.4, and 0.1. These parameters are modified to generate anomalous data.

A.2   Lenze Dataset

The Lenze dataset [24,41] is an industrial dataset that is gathered from a three-phase motor. The dataset comprises 524,289 data points and eight features, with the first six channels being related to the electric currents and voltages of the individual phases. We define physics-informed loss using the first six channels by leveraging Ohm's law, which establishes the relationship between voltage, current, and resistance in an electric circuit. The signals originating from the damaged bearing are classified as anomalies.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = R\frac{\mathrm{d}I}{\mathrm{d}t}$$

A.3   Electro-Mechanical Positioning System (EMPS)

This dataset contains three features with 24841 datapoints, respectively. The EMPS is a standard configuration of a drive system for prismatic joint of robots or machine tools [1]. In the Inverse Dynamic Model (IDM) of a robot, the joint torque or force τ is expressed as a function of the joint position q, velocity, and acceleration [1]. Here, anomalous data are synthesized by scaling the amplitude of the signals.

$$\tau = M\frac{\mathrm{d}^2 q}{\mathrm{d}t^2} + F_v\frac{\mathrm{d}q}{\mathrm{d}t} + F_c\,\mathrm{sign}(\frac{\mathrm{d}q}{\mathrm{d}t}) + \mathrm{offset}$$

A.4   Air Compressor Dataset

The air compressor dataset is an industrial dataset that comprises 261,640 data points and eight features: Volume, Pressure, Temperature, Mass, Mass Flow Rate, Volumetric Flow Rate, Flow Velocity, and Energy. We compute physics-informed loss based on the ideal law [34], which establishes the relationship between volume ($v = V/m$), pressure, and temperature.

$$P\frac{\mathrm{d}v}{\mathrm{d}t} + v\frac{\mathrm{d}P}{\mathrm{d}t} = R\frac{\mathrm{d}T}{\mathrm{d}t}$$

and the relationship between mass flow rate and volumetric flow rate $\dot{m} = \rho Q$. Here, leakage occurring in the air compressor system is considered an anomaly.

B   Model architecture and Hyperparameters

We constructed the diffusion model based on an encoder-decoder architecture. We used Long Short-Term Memory (LSTM) networks and SiLU activation function in the encoder-decoder network. The sliding window stride is 1; the number of LSTM layers and hidden neurons, sliding window size, learning rate, and batch size depend on the data set. All models were trained only on normal data. The data was scaled to the interval [-1, 1], and a 90:10 train-test split was used to evaluate for overfitting. Model training was performed using the Adam optimizer with a learning rate of 0.0001 and L2 regularization of $1e^{-6}$. For model training, diffusion time steps of T=100 are used.

The variances of the forward process are set according to a linear schedule [12] from $\sigma_1 = 10^{-4}$ to $\sigma_T = 0.05$. The variational autoencoder and autoencoder model constructed with the same architecture as the diffusion model. The hyperparameters of the trained models have been optimized in informal experiments, resulting in the values given in Table 5. The parameters of the PINN weight schedule are specified in Table 4

The physics-informed diffusion model and other physics-informed methods are trained with the same architecture and hyperparameter as the uninformed model for every dataset. The time derivative for the physics-informed neural network is computed using finite-difference approximations.

Table 4. PINN weight schedule

f(t)	m	n	l
Log-Sigmoid	0.01	0.1	0.1
Hard-Sigmoid	0.01	1	1
Sigmoid	0.01	1	1
ReLU	0.001	0.01	0.9

Table 5. Hyperparameters of all datasets

	EMPS	Predator-Prey	Lenze	Air Compressor
Batch size	128	128	256	128
Epochs	350	500	500	200
Sliding Window	150	100	250	200
c of [2]	$10e^{-3}$	$10e^{-3}$	$10e^{-3}$	$10e^{-3}$
Encoder Layers	3	3	3	3
Decoder Layers	3	3	3	3
Encoder Hidden Neurons	(3, 12) (12, 64) (64, 128)	(2, 8) (8, 16) (16, 32)	(8, 32) (32, 64) (64, 128)	(8, 32) (32, 64) (64, 128)
Decoder Hidden Neurons	(128, 64) (64, 12) (12, 3)	(32, 16) (16, 8) (8, 2)	(128, 64) (64, 32) (32, 8)	(128, 64) (64, 32) (32, 8)
Neurons in the latent space	128	32	128	128
K Means Clusters	8	4	16	12

B.1   Model Training and Inference Time

The training and inference times of the models are shown in Table 6. Here, inference time refers to the duration required by the model to detect anomalies in 1000 (700 normal and 300 anomalies) data points. The training time of the

model varies depending on factors such as the dataset size, window length, and number of training epochs. The inclusion of the informed loss function results in a moderate increase in training time, due to the added computation of the PINN loss. The autoencoder achieved the lowest inference time among all evaluated models. Due to the diffusion process, the inference speed of the diffusion model is slower than AE and VAE.

Table 6. Training and Inference Time of all datasets

	EMPS	Predator-Prey	Lenze	Air Compressor
Training Time				
DM	50.20 mins	5.25 hrs	1.02 day	4.42 hrs
TPIDM	52.87 mins	5.36 hrs	1.06 day	4.59 hrs
PIDM [2]	40.64 mins	5.55 hrs	1.06 day	4.33 hrs
PIDM [33]	67.2 mins	7.02 hrs	1.26 day	5.06 hrs
VAE	49.47 mins	4.05 hrs	1.30 day	5.03 hrs
AE	47.81 mins	1.91 hrs	0.47 day	3.65 hrs
K-MEANS	0.40 mins	0.03 hrs	0.19 day	0.041 hrs
Inference Time in seconds				
DM	1.7423	1.6084	4.7679	3.7639
TPIDM	1.7748	1.6106	4.8928	3.7801
PIDM [2]	1.7309	1.5100	5.1415	4.0328
PIDM [33]	1.8817	1.7797	5.9067	4.0531
VAE	0.0251	0.0188	0.0633	0.0528
AE	**0.0228**	**0.0137**	**0.0594**	**0.0464**
K-MEANS	1.3028	1.7321	2.6844	2.2074